

DOCUMENT RESUME

ED 177 936

MF 011 720

AUTHOR Kanaga, Kim
TITLE The Evaluation of Instruction.
PUB DATE Apr 78
NOTE 40p.; Not available in paper copy due to marginal legibility of original; Paper presented at the International Communication Association Conference (Chicago, Illinois, April 25-29, 1978)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS *College Instruction; *Educational Assessment; *Effective Teaching; *Evaluation Methods; *Faculty Evaluation; Higher Education; Instructional Improvement; *Multidimensional Scaling; Peer Evaluation; Self Evaluation; Student Evaluation of Teacher Performance

ABSTRACT

Some of the theoretical and methodological problems with current practices in evaluating instruction at the higher education level are reviewed. Controversy over the evaluation of instruction in higher education has resulted at least in part from inadequate instrumentation. The instruments for instructional rating now used include administrator evaluations, colleague evaluations, self-evaluations, and student evaluations. It is suggested that the teaching-learning relationship must be the focus of meaningful evaluations of instruction. A multidimensional scaling technique for precisely evaluating instruction that is based on a cognitive perspective of student learning is presented. The technique provides a means of assessing cognitive processes in a manner that has been found to be precise and reliable and enables the direct assessment of teaching effectiveness. A major advantage is that the proposed system does not require advance knowledge of all relevant criteria on which student responses are based, although specific attributes used for making distance judgments can be interpreted from the results. Making complete paired comparisons on all concepts that are provided allows multiple criteria to be simultaneously examined and reduces the influence of social desirability factors. This evaluation system also establishes explicit standards for direct evaluation; provides relevant information for administrators, instructors, and students; is precisely ratio-scaled; and permits powerful time-series analysis of the learning process, permitting instructors to monitor their effectiveness throughout the term. (SC)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

AUG 31 1978

The Evaluation of Instruction

by

Kim Kanaga

Department of Speech Communication

San Francisco State University

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

Kim
Kanaga

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

International Communication Association

Chicago, Illinois

April 25-29, 1978

THE EVALUATION OF INSTRUCTION

ABSTRACT

The evaluation of instruction has traditionally been a controversial issue among educators. The present paper argues that this situation is at least partially due to inadequate instrumentation. The paper examines some of the theoretical and methodological problems with current practices in evaluating instruction at the higher education level. It is argued that such evaluations should be based on a theory of the teaching-learning relationship. It is further suggested that a cognitive perspective of student learning would be most fruitful for the purpose of evaluating instruction. Such a perspective is then provided. On the basis of this perspective, a multidimensional scaling technique for precisely evaluating instruction is described. The paper concludes with a discussion of the advantages and limitations of the proposed procedures.

THE EVALUATION OF INSTRUCTION

Introduction

The evaluation of instruction has traditionally been a controversial issue among educators (Doyle, 1975; Fenker, 1975; Kent, 1966; Lewis, 1975; Miller, 1972). While most instructors would probably agree that there are potential advantages in carrying out such evaluations, Gray (1969) points out that they are usually less than enthusiastic about having their own classroom performance examined. This is particularly the case in higher education where the reasons for this lack of enthusiasm are quite evident. At the college level, very few instructors receive some sort of formal training (Doyle, 1975). In the absence of such training, instructors are left to either develop their own teaching skills or model their teaching after instructors from which they have had courses. The underlying assumption being that the instructor has, for certain courses, the necessary substantive knowledge and should be able to convey that information to students without much difficulty. The resistance to the evaluation of instruction may, then, be partially due to the instructors insight into this deficiency in their own instructional training.

Another issue contributing to this controversial situation is the current status of the reward structure within institutions of higher education. Retention, promotion, tenure, and pay raises at many major universities tend to be based more on scholarly producti-

vity than on instructional achievements (Astin and Lee, 1966; Doyle, 1975; Fenker, 1975; Hammond, Meyer, and Miller, 1969; Luthans, 1967). Even when teaching is taken into consideration, it is usually the quantity of students that is emphasized rather than the quality of instruction. Committee membership and other such responsibilities, in and of themselves, are usually not weighted heavily in the reward system. They do, however, involve working with peers and as such would be expected to elevate their importance to faculty members. Moreover, consulting and other outside activities typically provide substantial incentives. It would appear, then, that university systems are not structured to promote quality instruction. Instead, the emphasis on other responsibilities would seem to lower the priority placed on developing and maintaining superior teaching skills.

By their lack of enthusiasm, instructors may, in fact, be calling into question the practical utility of instructional evaluations (Rippey, 1975). In other words, instructors may see potential advantages to evaluating classroom performance, but have yet to see any of those advantages realized. Although administrators may now be at least looking at evaluation forms, they still appear to have little faith in them (Astin and Lee, 1966; Lewis, 1975). Furthermore, the information provided by present evaluation forms is of limited value to instructors who wish to use them for assistance in improving their teaching skills. Typically, those forms rate the instructor on various criteria (Genova, Madoff, Chin, and Thomas, 1976). The instructors are then informed as to how they rate in comparison with other instructors. However, there is no indication whether the individual instructor's position is more or less effective than the other instructors

but merely how much they differ from other instructors. Even if specific strengths and weaknesses were precisely pinpointed, these forms rarely contain any recommendations for how specific teaching skills may be improved. Thus, the skepticism surrounding instructional evaluation may be, in part, due to the failure of current evaluation procedures to produce any meaningful results.

This latter point suggests that at least some of the problems associated with instructional evaluations may be found in the instrumentation rather than the lack of concern among instructors and administrators. An extensive survey conducted by the American Council on Education (Astin and Lee, 1966) revealed that systematic measuring instruments are, in fact, not a widely used means of evaluating classroom performance. Instead, they found that by far the most frequent sources of information are the subjective judgments of department chairpersons (85.1%) and deans (82.3%) followed distantly by scholarly productivity (43.8%) and in informal opinions of colleagues (48.9%) and students (41.2%). As Gustand (1961) put it, ". . . the most frequently cited sources of information can be described simply as hearsay."

The extensive use of such imprecise procedures may, very well, be an indication of the quality of the measuring instruments that are currently available. Doyle (1975) and other contemporary educational researchers (Blackburn and Clark, 1975; Kauffman, Hallahan, Payne, and Ball, 1973; Lewis, 1975; and Miller, 1972) have suggested that current evaluation procedures are inadequate. This conclusion was also reached by the Commission on Academic Tenure in Higher Education (1973). In their report, they recommend that "professional associations and research centers continue and expand their efforts to develop more

sophisticated conceptions of teaching effectiveness and more reliable methods of evaluating teaching . . . ". This paper will be such an undertaking. The remainder of the paper will: 1) critically examine the current procedures for evaluating instruction, 2) propose more precise procedures for evaluating instruction, and 3) discuss the advantages and limitations of the proposed procedures for improving the existing educational process.

Current Practices in the Evaluation of Instruction

There are probably as many different instructional rating instruments as there are colleges and universities which employ them. Each one purports to tap the most essential characteristics of teaching effectiveness. It would follow that if these instruments are, in fact, measuring the same phenomena, there should be some commonality among them. While some of the instruments are similar in appearance and claim to be measuring the same dimensions (Doyle, 1976), the items used to tap these dimensions are quite different (Genova et al., 1976). Since the items are not comparable from instrument to instrument, it is quite possible that even if they were examining the same dimensions, they may be focusing on a different range of those dimensions. Still other instruments are quite diversified in terms of the source and type of information they obtain. Consequently, a meaningful discussion of current evaluating instruments as a whole is questionable. Instead, these instruments can be more clearly examined within the context of the following four categories: administrator evaluation, colleague evaluation, self evaluation, and student evaluation.

Administrator Evaluation

In most cases, it is ultimately the administrators responsibility to evaluate the teaching performances of faculty members. The administrator is typically asked to make an overall assessment of teaching effectiveness as well as of other faculty duties. While more specific information is probably desirable, the more crucial issue in administrator evaluation is to determine on what basis such general assessments are made. Most administrators are unable to actually observe the classroom behavior of their faculty. The reason for this boils down to a simple matter of priorities. Although the evaluation of faculty teaching is an administrative concern, it is only a portion of the responsibilities delegated to a dean or department chairperson. Consequently, administrators do not have the time necessary to thoroughly evaluate each faculty member. They must, instead, rely on other sources of information.

Even if time permitted, the utility of an overall assessment based on the observations of any single individual is questionable. That is, direct observational ratings of this sort are based on individual perceptions which are at least partially determined by past experiences. While explicitly stated criteria for making ratings may minimize this potential source of bias, it is difficult, in the absence of other information, to determine if the criteria have been appropriately applied by any one individual. Several possible problems are noteworthy. Administrators faced with such a task may be partial to certain teaching styles which would not necessarily render other methods as less effective but would merely be an indication of personal preferences. It could also be difficult for them to determine the extent to which the

instructor is facilitating or impeding student learning of course materials. What is clear and understandable to the administrator may be vague and incomprehensible to the students. What may seem as disorganized to the administrator may instead provide the students with the flexibility needed to get them more involved in class projects. Another potential source of bias may result from a difficulty in assessing teaching performance independently of other components which comprise the faculty members' overall contribution to the department and/or college (Blackburn and Clark, 1975). Moreover, these and other individual biases could result in a lack of uniformly applied standards which, in turn, may result in serious morale and/or legal problems.

This is not to say that these biases will occur or that they are by any means intentional. But even if administrators were aware of their biases, they may not be able to eliminate them and in attempting to do so may cause biases in other directions. It would appear, then, to be to the administrators' advantage to seek out the most precise, reliable, and valid information at their disposal and to utilize that information in evaluating teaching effectiveness.

Colleague Evaluation

Colleague evaluations tend to parallel administrator evaluations in both purpose and problems. These evaluations are primarily used to provide administrators with information regarding the performances of faculty members. While some peer rating forms (Genova et al., 1976) ask for summary information, others (Blackburn and Clark, 1971; Miller, 1972) allow colleagues to make more specific assessments and as such these instruments may also be used to assist instructors in

improving their teaching skills.

There are several problems which accompany this type of evaluation. Faculty members are not immune to the biases that administrators are subject to when directly observing teaching behavior. In fact, they are probably more prone to bias. Faculty members have more personal contact with one another and would themselves be in a position to subsequently be evaluated by those faculty members whom they must evaluate. However, Genova et al. (1976) point out that faculty members rarely visit each others' classroom for the explicit purpose of assessing teaching performance. As a result, their evaluations must also be based on other sources of information.

A more serious issue resulting from peer evaluations is the impact that they may have on the faculty. Eble (1970) as well as others (Fenker, 1975; Miller, 1972) have found strong opposition to colleague evaluations. Fenker (1975) reported that resistance was so strong that his attempt to collect colleague evaluations had to be abandoned. This suggests that, for at least administrative purposes, the cost of such information may outweigh the benefits.

Self Evaluation

As evidenced by the lack of attention in the literature, formal self evaluations are relatively rare in higher education. Moreover, the research that is available (Blackburn and Clark, 1971; Centra, 1972) tends to suggest that instructors inflate the assessments of their own teaching skills. If, on the other hand, faculty members were able to accurately assess their own effectiveness as a teacher, it is questionable whether individuals should be put in the position of

providing the information that would warrant their own dismissal or would call for a promotion or pay raise. It may be desirable to think that professionals could and would make such selfjudgments. To make such an assumption is, however, unrealistic and would establish a potentially detrimental reward system.

Formal self evaluation may however, be used as a means of improving teaching skills. Having instructors seriously reflect on their teaching behaviors could be quite helpful to some instructors; particularly those interested in improving their teaching ability but not knowing how to proceed. Self evaluation could provide them with some needed direction. Under these conditions, additional information from other sources would also be desirable.

Student Evaluation

Student rating forms are by far the most widely used means of formally gathering evaluative information for each of the purposes previously mentioned (Doyle, 1975). That is, student evaluations have been used to: 1) provide information for administrative purposes, 2) provide diagnostic feedback for instructors, 3) provide information for students and their advisers, and 4) provide data for educational researchers.

The most profound issue regarding student appraisals of teaching effectiveness is a debate over the students' ability or competence in making such judgments. While administrator and colleague evaluations involved assessments by individuals with at least comparable experiences and training, the opponents of this type of evaluation argue that students are not qualified to make such assessments (Bryant, 1967;

Hildebrand, 1972). Supporters of student evaluations counter by arguing that as the consumer of the instructor's efforts, students are in the best position to actually know the effectiveness of the instructor (Astin and Lee, 1967; Grush and Costin, 1975). The issue is, however, contaminated by the limitations of the evaluation procedures that are presently in use.

Most educators would probably agree that student responses based on learning is desirable. But current instruments fail to actually measure student learning as a criterion for determining teaching effectiveness. Instead, they typically require students to indicate, on some sort of forced choice scale, the extent to which a given statements accurately characterizes a specific instructor. These are intended to reflect a various underlying components which, when taken as a whole, constitutes teaching effectiveness. Students are asked to make ratings on such characteristics as the instructors' flexibility, motivation, rapport, teaching style, and interaction patterns (Doyle, 1975). Extensive reviews of empirical research by Doyle (1975) and Costin et al. (1972) show rather convincingly that students are quite capable of making such assessments. Although these results are still occasionally questioned, it becomes more crucial to determine the extent to which the components are related to the actual learning that takes place in the classroom. In the absence of direct measures, it is assumed that, collectively, the dimensions tapped by the measuring instrument are a reflection of the teaching-learning process. The issue, then, becomes as Rippey (1975) points out, not a question of whether or not students should be used for the purpose of determining teaching effectiveness, but rather a question of how to best utilize this potentially valuable source of information.

In addition to the problems created by the nature of the information that present evaluation procedures obtain from students, they also suffer from several methodological shortcomings. All of the instructional evaluation instruments that are either reproduced or described in the literature are subject to at least one of the following limitations:

- 1) All relevant criteria which prospective respondents use to evaluate instruction must be known in advance and/or appropriately captured in the measuring instrument. Statements characterizing specific instructional qualities must reflect these criteria. The lack of agreement regarding what are the essential criteria and how they are to be tapped suggests that current instruments are not measuring all of the criteria by which instruction is evaluated.
- 2) Respondents are typically required to assess the instructor separately for each of the criteria that are employed in the instrument. Since these criteria are seldom independent of one another, obtaining additional information regarding the interrelationships among these criteria would provide a more complete picture of the evaluation process.
- 3) There is a general lack of precision of measurement in instructional evaluation forms. Those instruments extensively utilize some sort of Likert and/or semantic differential type scales which, at best, can be considered ordinal measure. The necessary precision of measurement is dependent on the phenomenon under consideration. Ambiguous conceptualizations of teaching effectiveness makes this issue unclear. However, unless evidence is available to suggest that ordinal scales would be adequate, more precise measures should be sought.
- 4) Instructional evaluation instruments are rarely designed with over-time analysis in mind. Virtually all of the instruments that are in use are set up as ex-post-facto measures. If the teaching-learning relationship is a dynamic one, which few educators would dispute, it should be examined across time. End of the term measures only provide a static and potentially misleading description of a complex process. Moreover, they come too late for instructional modifications to be made.

- 5) Criteria that are explicitly laid out in the form of statements concerning specific instructional qualities runs the risk of being quite easily distorted by respondents. Administrators, colleagues, and students all have vested interests in the effectiveness of the instructors they are asked to evaluate. Consequently, the ability to bias their responses should always be minimized.
- 6) In order to determine just how effective an instructor is, what constitutes effective instruction must first be identified. While much attention has been focused on what criteria ought to be incorporated into an instrument, there is usually no provision made for determining some optimal level of performance for each of these criteria. Without an explicit standard for comparison, interpretation of any results would be at least somewhat questionable.

In summary, it has been pointed out that assessments of teaching effectiveness have been used to provide information for administrative purposes, for the improvement of teaching skills, for educational research, and for counseling students. As sources of information, it was suggested that administrators and colleagues are unable to invest the time necessary to thoroughly evaluate teaching effectiveness and are highly susceptible to a number of biases. Self evaluation was found to be impractical. Students appear to be the most promising source of information because they have much more exposure to the instructor in the classroom setting, they are recipients of the teaching process, and their numbers are usually of sufficient size to minimize error resulting from individual differences. However, until a precise means of measuring student learning is developed, their potential value will not be fully realized.

It seems quite clear, then, that if the formal evaluation of instruction is to play a vital role in the educational process, the teaching-learning relationship must be the focus of theoretical development and a measurement system not suffering from the limitations

of current procedures must be devised.

Student Learning as a Criterion for Determining Teaching Effectiveness

In pointing out that "the ultimate criterion of good teaching is student learning," McKeachie, Lin, and Mann (1971) are not making any startling revelation. They are merely stating what educators have known for a long time. Learning is, after all, the purpose for teaching (Tyler, 1958; Cohen and Brawer, 1969). It has, however, rarely been used as a determinant of teaching effectiveness because traditional measures of learning are imprecise (Kauffman, Hallahan, Payne, and Ball, 1973) and not conveniently applied to the educational environment. Before turning attention to the theoretical under-pinnings of learning measures, it would be informative to first discuss the means that have been used to measure student learning when such attempts are made. One means of getting at learning has been to tap into the grading system (Costin, *et al.*, 1972). It intuitively makes sense to think that grades differentiate levels of student learning. But, at best, this is only an indication of the effectiveness of an instructor with various students. Using grades to determine overall teaching ability only confuses effectiveness with the instructors evaluation of student performance. That is, a final exam in which all students correctly respond to all items does not necessarily mean that the instructor was maximally effective. It would probably suggest that the instructor was unable to construct a test that adequately discriminated between various levels of student learning. Conversely, an instructor assigning abnormally low grades may be very effective but rigorous. Standardized achievement tests have also been used to determine student

learning (Gessner, 1973). However, they often lack flexibility and are not available for all subject matters. The most rarely used means of measuring student learning has been to merely ask students what or how much they have learned. Even though researchers (Doyle and Whitely, 1974; Frey, 1973) have found that students' responses to these types of questions are positively related to final exam scores, their accuracy in making such direct judgments is somewhat questionable. Moreover, their responses are quite easily distorted. Although these latter two types of measures have shown some promise, they are still not without serious limitations. This lack of adequate measures of classroom learning can be traced to the theories from which the measurement systems were derived.

Learning Theory

The most influential theories of learning (Bolles, 1975; Hilgard and Bower, 1975; Pittenger and Gooding, 1971; Snelbecker, 1974) have been behaviorally based in the traditions of Thorndike, Pavlov, and Skinner. These theories generally define learning as a relatively stable change in a behavioral pattern resulting from repeated exposure to some external stimuli (Snelbecker, 1974). They were comprehensively designed to encompass learning in all animals and as such were specifically tested on rats, pigeons, and monkeys with the express purpose of generalizing the results to humans. That is, it was felt that if other animals adhere to these principals of learning, humans, despite their cognitive potential, must also be subject to them.

To use such a theoretical framework in formal education would require specific behavioral outcomes for all course material to be identified and measured. While this may be a desirable goal, an

evaluation system based on this information must either assume comparability of outcomes and measures, or make some allowances for differences. Neither alternative is desirable. A more serious problem has to do with the incompatibility of the behavioralist learning theory and the typical way courses proceed. Course material is usually covered continuously by topics. Once taught, topics are not frequently repeated. Consequently, the repeated exposure to stimuli, which is usually considered to be a necessary condition for learning, is quite often not present. Moreover, behavioral learning theories typically call for the response to follow immediately after exposure to the stimulus. In the classroom situation, it is not uncommon to have weeks go by between the time certain course material is covered in class and the time that students are expected to utilize that information. What happens to the information during this time between stimulus and eventual response is not easily accounted for by the strictly behavioral perspective. It seems clear, then, that while these theories may be applicable in other contexts, they are of limited utility in evaluating classroom instruction.

A more appropriate theoretical orientation can begin to be developed by first focusing attention on the work of Hebb (1958, 1960) and Osgood (1957, 1963). Hebb (1958) contends that there exists thought or mental processes "which themselves independent of immediate sensory input, collaborate with that input to determine which of the various possible responses will be made and when." In other words, these mental processes mediate responses to incoming stimuli. What is being suggested here is essentially a two-stage activity. Stage one involves the internal interpretation of external stimuli.

In stage two, the interpretation serves as an internal stimuli for a given response. This internal interpretation is what Osgood (1957) refers to as "meaning" which is gained from an association made between specific stimuli and the responses which they have elicited. It is further suggested that because certain stimuli are linked to similar responses, these stimuli also become associated with one another. New stimuli can then be learned by being associated with other stimuli that already have specific responses attached to them.

In a study investigating the associations between various stimuli, Noble (1952) found the existence of chain associations. That is, one stimulus is associated to another which in turn is associated with another and so on. In other words, the internal response to an external stimulus can serve as an internal stimulus for yet another internal response and this process can go on indefinitely or until the ultimate external response is made. Theoretically, if the chain were long enough, the resultant external response need only be remotely related to the initial stimulus. Nevertheless, this indicates that rather than internally possessing an enormous number of unrelated associations, stimuli are probably incorporated into an overall structure and are at least indirectly related to one another.

This idea of an internal structure appears to be quite consistent with the position taken by Tolman (1932) and the Gestalt theorists (Koffke, 1935; Kohler, 1947; Wertheimer, 1959). From a more cognitive perspective, they viewed learning as the acquisition and organization of information about the environment. In other words, their emphasis was on "man's ability to see meaningful relationships and structure (Snelbecker, 1974). Tolman (1932) did not, however,

abandon behaviorism. Instead, he also theorized that cognitive processes intervened between external stimuli and overt behavior.

Stimulus information is incorporated into an individual's cognitive structure until such a time as it is called into play by some activity. Thus, the relationship between external stimuli and behavioral response is indirect and not necessarily immediate.

In order for the cognitive structure to become useful in the assessment of student learning, it must first be determined how the structure is organized. What is the basis by which associations are made? What is the criteria for relating stimuli to one another in the cognitive structure? Bruner (1966) suggests that stimuli are categorized as concepts or objects of an individual's experience. Objects are associated with one another when they belong to the same category. Categories are established according to certain essential distinguishing characteristics or what Bruner refers to as criterial attributes. "To categorize is to render discriminately different things equivalent, to group the objects and events and people around us into classes, and to respond to them in terms of their class membership rather than in terms of their uniqueness" (Bruner, 1966). Thus, learning involves knowing: 1) what attributes are available, 2) when they should be applied, and 3) how they should be applied. Successful learning occurs when stimuli or objects are appropriately categorized.

In the absence of more specific information, a categorical system may be sufficient to explain the organizational pattern of the cognitive structure. However, finer discriminations are usually possible. Individuals are able to identify and respond to unique entities in the environment. For example, the major means of transportation that is

usually found in the garage is not just a car, but a given make and model produced in a certain year by a specific automobile manufacturer. The use of additional properties or attributes such as color, rust spots, dents, mileage, and license plate numbers would eventually identify a single car. Maintenance and use of that car is likely to depend not only on the categories that it belongs to but also on its idiosyncracies. In other words, as more information becomes available, the categories in which an object can be placed become narrower and narrower until such a time as the object becomes a category in and of itself. The object is then differentiated from everything else and can be responded to according to its unique and/or categorical characteristics. Thus, objects gain their meaning by their placement on attributes relative to other objects.

Woelfel (1974) has further suggested that the information extracted from the environment is organized into a pattern of similarities and differences among the concepts that are symbolized by words in the vernacular language. Each concept is uniquely defined by its interrelationships with all other concepts. Changes in the meaning of concepts are cognitive processes and result in modifications of the overall cognitive structure. For example, speech, as an academic discipline which serves as a categorical domain of meaning, is defined according to how it is differentiated from art, biology, business administration, history, physical education, psychology, and all the other academic disciplines. While, at one time, speech was strongly identified with the humanities, it has begun to be more strongly associated with the social sciences. As a result, the cognitive structure of those in the academic community that are aware of the changing

emphasis in speech-communication departments would be expected to be experiencing a reorganization in which the location of speech within the structure is undergoing change.

Given this foundation of cognitive processes, learning can be defined as the evolution of meaning. This definition portrays learning as a continuous process where the meanings for concepts are initially created by setting observations into correspondence with specific symbols and are subsequently molded by the accumulation of concept-relevant information. In other words, learning involves the development and modification of an individual's attitudes, beliefs, and values.

In terms of formal education, learning must be considered in conjunction with teaching. The function of teaching is typically to develop consensually shared meanings for the concepts that define the domain of instruction. In contrast, teaching is also occasionally used to broaden student perspectives. Whether the intent is to develop individual or collective meanings, all courses are usually set up around some overriding theme or goal. The relationships between the concepts that make up this theme constitute the basic foundation on which the course rests. Within this framework, specific course objectives may be dealt with by a number of different teaching techniques. The success of students in meeting these objectives is, of course, important. Just as important, if not more so, is the extent to which the students have properly internalized the meanings of the basic concepts on which the course is based. In this respect, teaching effectiveness can be seen as the progress of the students, as an aggregate, in learning the interrelationships between the concepts which define the overall structure of the course.

A potential limitation of a cognitive perspective is its generalizability to all learning situations. In particular, the utility of an evaluation system based on a cognitive conceptualization of learning may be called into question in skill or performance oriented courses. The primary objective of a calculus course, for example, would probably be to teach students to actually calculate differentials and integrals. An evaluation based on the student's ability to make these calculations would be a more direct and probably more desirable measure of teaching effectiveness. On the other hand, a good grasp of the course material should also be reflected in the development of specific meanings for these operations. The distinction appears to be one of knowing how to perform a specific activity versus understanding when it is appropriately performed. Both need to be learned and neither is precluded when cognitive processes are considered to intervene between external stimuli and resulting behaviors.

It should also be explicated pointed out that this theoretical framework is not strictly cognitive. That is, when students provide information regarding their cognitive structures, they are in essence behaviorally responding and it is assumed that those responses reflect the classroom stimuli that they have been exposed to. Rather than focusing on ultimate behavioral outcomes which vary from course to course, the emphasis is on the mediating cognitive processes which it is assumed that all students must use to interpret and assign meaning to stimuli. The result is that the behavioral responses required of students, regardless of their courses, is completely standardized. Moreover, it is assumed that the cognitive structure serves as the basis upon which other behaviors are made.

Measurement of Teaching Effectiveness

One of the principle objectives to early cognitive oriented theories of learning was the lack of adequate measurement of cognitive processes (Snelbecker, 1974). It was assumed that since cognitions are unobservable, they could not be measured with any degree of success. Ironically, this argument can now be turned around and used in support of a cognitive perspective of learning. It is not that cognitive processes are now observable but rather that useful procedures have been devised which allow these processes to be accurately measured. Metric multidimensional scaling provides a means of assessing cognitive processes in a manner that has been found to be precise and reliable (Gillham and Woelfel, 1976; Woelfel, 1977). Moreover, a multidimensional analysis as is to be described here satisfies all of the limitations of measuring teaching effectiveness that are found in traditional evaluation instruments.

Multidimensional scaling is based on the fundamental concept of psychological distance (Helm, Messick, and Tucker, 1959). That is, the perception of difference between stimuli or objects is the basis of the measurement scheme. Woelfel (1972) suggests that differences "among objects (whatever those objects may be) may be represented by a continuous numbering system such that two objects considered to be completely identical are assigned a paired dissimilarity score or distance score of zero (0), and objects of increasing dissimilarity are represented by numbers of increasing value." To utilize the set of real numbers in such a manner, a rule must be established for setting the numbering system into correspondence with perceptions of difference.

Stevens (1951), Torgerson (1958), and others (Campbell, 1928; Suppes and Zines, 1963) have stated that a rule for quantifying distance or difference must stipulate an arbitrary standard difference to which all other differences are to be compared. Such a rule is provided by Einstein (1961):

For this purpose (the measurement of distance) we require a 'distance' (Rod S) which is to be used once and for all, and which we employ as a standard measure. If, now, A and B are two points on a rigid body, we can construct the line joining them according to the rules of geometry; then, starting from A, we can mark off the distance S time after time until we reach B. The number of these operations required is the numerical measure of the Distance A B. This is the basis of all measurement of length.

To use Einstein's rule for measuring the perceived differences among objects, all that needs to be done is to arbitrarily stipulate that the difference between any two objects is some designated distance and that this distance is the standard of comparison for all other pairs of objects. This can be accomplished by wording a question in the following form:

If the difference between a and b is u units, how different are x and y?

Responses to this type of question would be ratio measures of the perceived differences between pairs of objects.

When this ratio rule is applied to N concepts, $N(N-1)/2$ non-redundant paired comparisons are possible. Completing all such pairs produces a $N \times N$ symmetric matrix D for an N concept domain. This matrix, then, represents the overall pattern of differences among the concepts in the domain. Woelfel (1972) further states that:

. . . the definition of an object or concept is constituted by the pattern of its relationship to other objects, the definition of any object may be represented by an $1 \times n$ vector, $d_{11}, d_{12}, d_{13}, \dots, d_{1n}$, where d_{11} represents

the distance or dissimilarity of object 1 from itself thus $d_{11}=0$ by definition, d_{12} represents the distance or dissimilarity between objects 1 and 2, and d_{1n} represents the distance between the 1st and nth objects. Similarly, the second object may be represented by a second vector, $d_{21}, d_{22}, d_{23}, \dots, d_{2n}$, and the definition of any set of concepts or objects may therefore be represented in terms of the matrix

$$\begin{matrix} d_{11}, d_{12}, \dots, d_{1n} \\ d_{21}, d_{22}, \dots, d_{2n} \\ \cdot \quad \cdot \quad \quad \cdot \\ \cdot \quad \cdot \quad \quad \cdot \\ d_{n1}, d_{n2}, \dots, d_{nn} \end{matrix}$$

where any entry d_{ij} represents the dissimilarity or distance between i and j .

Since the primary concern in evaluating teaching effectiveness is with aggregates of students, potential unreliability resulting from individual differences in responses is minimized by averaging dissimilarity scores across the students in a specific course. This procedure yields a means distance matrix D which represents the average dissimilarity among concepts which constitute the domain of instruction. This matrix is transformed into a centroid scalar products matrix B (Young and Householder, 1939) which when factored (Jacobi, 1846) results in a cartesian coordinate system of orthogonal dimensions or axes. A rectangular matrix F is then constructed with the dimensions as columns of the matrix and with the rows representing the projections of the concepts on each of the dimensions. These procedures, which are described in much more detail by Torgerson (1958), Woelfel (1972, 1974, 1977), Woelfel and Danes (1977), and Serota (1974), essentially map the structure of the domain.

Cognitive processes such as learning can be represented by comparing a series of these spatial configurations gathered at several points in time. The coordinate systems generated at each point in time can be rotated and translated into a least square best fit with one another (Woelfel, 1977; Woelfel and Saltiel, 1974). In other words, this procedure is used to establish a common frame of reference from which changes in the meanings of objects or concepts can be observed. Changes in the spatial location of the concepts over time is then interpreted as motion through the space and as such can be mathematically expressed as velocities and over multiple time periods as accelerations. This enables student learning to be precisely assessed over the length of instruction.

The implementation of this multidimensional scaling technique for the purposes of evaluating teaching effectiveness is straight forward. What first needs to be done is to identify the relevant concepts that will be scaled into the space. Since these concepts define a particular domain of meaning, they must be topic or course specific. Moreover, these concepts must describe the basic nature of the course, not evaluate it. In other words, they must constitute the overall framework on which the entire course is based. Such concepts typically make up the general course objectives or outline. For example, one of the primary goals of an introductory communication course is to provide students with some sort of definition of communication and a notion of how the study of human communication can provide them with some useful information. From this objective, the following key concepts can be extracted: communication, process, information, valuable, meaning, social science, humanities, and me (representing

the self-concept, see Woelfel and Danes, 1977). A second goal is typically to differentiate several types of communication systems and to bring out the function of communication within each system. Depending, of course, on the systems that are explicated, the following concepts might emerge: communication, information processing, interpersonal relationships, group interaction, organizations, mass media, social change, persuasion, meaning, leadership, self-concept development, efficiency, and socialization. A course in research methods, which covers more technical information could be described by concepts such as: science, theory, measurement, percision, mathematics, function, causality, experimental design, reliability, and validity.

A predictable pattern of interrelationships among the selected oncepts should develop over the duration of the course. Assuming the course materials were being appropriately taught, the meanings given these concepts by the students would be expected to begin to converge on the projected meanings provided by the course objectives. Going back to the example of the introductory communication course, several changes would be expected to occur. Leadership, for instance, is not intuitively thought of as a communication construct. If it is effectively taught as such, a stronger association between communication and leadership would be expected to emerge. Assuming a strong relationship between cognitive processes (i.e. attitudes and beliefs) and behaviors (see Liska, 1975; Cushman, 1977) the subsequent communicative behaviors of the students should become more salient to them when they are placed in a position of leadership. If the course objectives are attained, the students would also have begun to more fully realize the value of communication. Since concepts associated

together have been found to converge on each other in the multi-dimensional spatial configuration (Woelfel, Cody, Gillham, and Holmes, 1976), the perceived dissimilarity between communication and valuable should decrease.

Doyle (1975) has suggested that a useful criteria for evaluating instruction is the stimulation of student interest in the subject matter of the courses that they are taking. This can be directly assessed for all courses by examining the relationships between the concepts which constitute the instructional domain and the concept "me" which as previously mentioned represents the students' self-concept. In particular, student interest is reflected in the relationship between the self-concept and the course title or some facsimile of it. Increased student interest then would be expected to result in the students' identifying more closely with the concept which encompasses the course's subject matter.

To determine the extent of teaching effectiveness, more specific information is required. It is not enough to merely examine general trends in student learning patterns. A standard for comparison of student progress towards learning appropriate interrelationships among the selected concepts must also be provided. Two primary sources are available for the construction of such a standard. Already knowing the content of the course and its goals, instructors would be in a good position to provide this information. The instructors own meaning for the specific domain of instruction may not be the same as that which would be expected of students. Instructors could instead provide a multidimensional configuration based on their expectations of the responses that could reasonably be expected of students who

have mastered the course material. Unless, however, a sizeable group of qualified instructor could be used for this purpose, the procedure could result in substantial errors due to inaccurate projections.

The second source of information is students. That is, the meanings developed by students who have previously been successful in learning specific course material could provide a standard for comparing other students. There may, however, be problems in identifying the appropriate students and in comparing levels of student achievement across time. It would seem, then, that the most desirable standard of comparison would be a weighted average between instructors expectations and previous successful students. This would minimize the extent of errors in instructor estimates and would contain the flexibility for future student performances to exceed or differ from current levels. The weights are dependent upon the number of instructors and students contributing to the standard, changes in the course, and differences in students enrolling in the course.

A standard representing desired course outcomes enables student learning to be more thoroughly assessed. Periodic comparisons of student responses with the standard reveals not only what has been learned but also what needs to be learned. The extent to which students progress toward this standard is, then, a direct measure of teaching effectiveness.

While the procedures described thus far, are more than adequate to precisely assess teaching effectiveness, a thorough evaluation of instruction should also contain more detailed information regarding the specific instructors' classroom behavior. In addition to knowing the extent of student learning that took place in a given course, it

is also informative to know what the instructor did or did not do that resulted in the learning that occurred. Multidimensional scaling has been found to be a useful tool in this type of situation as well (Cody, 1976; Wakshlag and Edison, 1975).

For this purpose, salient teaching characteristics need to be identified and scaled into the multidimensional space along with the concept "instructor" which, of course represents the specific instructor whose teaching performance is being examined. A list of appropriate concepts of characteristics (i.e. clarity, stimulating, effective, rapport) should be obtained from students, instructors, administrators, and from previous research (Dashpande, Webb, and Marks, 1970; Sharon, 1970). These concepts must then be submitted to a Q-sort technique such as the one described by Wotruba and Wright (1975). Essentially what needs to be done here is to locate those concepts which are perceived to be the most important and which students are able to accurately make judgments of. Only concepts that meet both of these criteria as rated by students, instructors, and administrators should be included in the instrument. The lack of consensus would likely result in the obtained information not being fully utilized by all those who could benefit from it.

Along with the obtained concepts and "instructor" a concept entitled "the ideal instructor" could also be used for comparative purposes. The ideal instructor would be described as an individual from which the students perceive that they would maximize their learning potential. In other words, the descriptive teaching characteristics would be used to define an optimally effective instructor. The actual instructor would be compared to the ideal one according to more or less

dissimilarity with these concepts. Since any direct comparison between the ideal and actual instructor by the students would be highly susceptible to bias, evaluation instruments should contain one or the other. The comparison can, then, be made more impartially. The information obtained from this procedure could be used to assist instructors in identifying particular teaching skills that need improvement and to assist students in locating instructors with specific desirable teaching styles.

Advantages

The overall advantage of the measurement system presented here is that it does not suffer from the problems confronting traditional evaluation procedures. Specifically, this system of evaluating instruction 1) is based on a theory of the teaching-learning relationship and as such enables the direct assessment of teaching effectiveness. It 2) does not require advance knowledge of all relevant criteria on which student responses are based. This measurement technique assumes that the overall perception of difference precedes the perception of attributes. Consequently, specific attributes used for making distance judgments need not be known nor directly utilized in the measuring instrument. They can, however, be subsequently interpreted from the results. Making complete paired comparisons on all concepts that are provided 3) allows multiple criteria to be simultaneously examined. Moreover, the paired comparison technique 4) is not readily susceptible to the influences of social desirability factors. Doyle (1975) reports that several studies, using traditional Likert type measures, have found that students become more lenient in their evaluations when they

are told that the results are to be used for administrative purposes. This is less likely to occur in this situation because actual comparisons to the standard of evaluation are not directly made by students and the scaling procedure provides no inherent indication of how one would go about making an instructor appear more favorably. This evaluation system also 5) established explicit standards for direct evaluation; 6) provides relevant information for administrators, instructors, and students; 7) is precisely ratio scaled; and 8) permits powerful time-series analysis of the learning process. This last advantage is particularly important because it enables instructors to monitor their effectiveness throughout the term. As a result, appropriate modifications in lesson plans and teaching styles can be made.

Limitations

This evaluation system is, however, faced with two potential problems that should be mentioned. One objection to the system has to do with the burden that it places on the students. Multiple measures for several courses could begin to require students to spend a good deal of their time evaluating instruction. Several steps could be taken to minimize this problem. First, samples of students could be taken at each time interval so that all students would not always be asked to participate. Secondly, the instrument could be broken up such that students would only have to fill out a portion of the total number of responses. Finally, specific relational patterns may begin to reappear with a great deal of regularity after the instrument has been in use for awhile. It would only be necessary to periodically check to

make sure that these patterns have not been broken. Consequently the number of required responses would be reduced. Although these remedies help alleviate this problem, they do not eliminate it. Active student participation is necessary to make this system successful. It is expected, however, that their efforts will be rewarded with generally improved teaching and with more detailed information regarding the teaching styles of perspective instructors.

This evaluation system may be difficult to adapt to skill and performance oriented courses. However, the comparison of the actual instructor to the ideal instructor would still provide useful information. Moreover, it would also be of interest to determine certain attitudes such as favorability toward the course's subject matter. After all, training students to be excellent typists would not reflect teaching excellence if as a result, all of the students hated to type and would avoid doing so in the future. It should also be pointed out that the skill and performance oriented courses are in much less need of new evaluation procedures. That is, the extent to which students master the skills taught has always provided a directly observable means of measuring teaching effectiveness.

Future Research

The immediate concern with research is to thoroughly examine the reliability and validity of the evaluation system. Reliability can be measured by simply including several paired-comparisons more than once in the instrument. Concepts whose meanings are not expected to change could be scaled into the space in order to examine the stability of the system over time. Control groups could also be used

for comparative purposes. Validity can be addressed by partialling students according to the grade they received in the course. A strong positive relationship would be expected between the grade received by the students and their success in learning the appropriate meanings for the domain of instruction. Moreover, a negative relationship should exist between student grades and the perceived difference between the actual and ideal instructors.

Beyond this preliminary work, the research possibilities for this system are enormous. The teaching-learning relationship is the foundation of formal education. A precise, reliable, and valid measure of teaching effectiveness which is based on this relationship can be very useful to educational researchers. One specific application that should be mentioned is the use of this system to assist instructors in improving their teaching skills. The use of specific teaching characteristics in the multidimensional analysis enables areas that need improvement to be located. In addition, the effectiveness of various teaching styles (made up from combinations of these characteristics) can be found. By additionally video taping and analyzing a sample of instructors, specific behavioral indicators for each characteristic and style can be identified. Training programs could then be devised from this information.

Summary and Conclusions

This paper reviews some of the theoretical and methodological problems with current practices in evaluating instruction at the higher education level. It is argued that such evaluations should be based on a theory of the teaching-learning relationship. It is further suggested that a cognitive perspective of student learning would be the

most fruitful for the purpose of evaluating instruction. Such a perspective is then provided. On the basis of this perspective, a multidimensional scaling technique for precisely evaluating instruction is presented. The paper concluded with a discussion of the advantages and limitations of the purposed technique.

The purpose of this paper was to explore some current problems and suggest a possible solution. It was not intended to be definitive in nature. A substantial amount of research will have to be done before anything conclusive can be said about the evaluation system proposed here.

REFERENCES

- Astin, A. W., Lee, C. B.T. Current practices in the evaluation and training of college teachers. *Educational Record*, 1966, 47, 361-364.
- Blackburn, R.T., Clark, M. An assessment of faculty performance: Some correlates between administrator, colleague, student and self ratings. *Sociology of Education*, 1975, 242-256.
- Bolles, R. C. *Learning theory*. New York: Holt, Rinehart, and Winston, 1975.
- Bruner, J.S. *Toward a theory of instruction*. Cambridge: Harvard University Press, 1966.
- Bryant, P.T. By their fruits ye shall know them. *Journal of Education*, 1967, 38, 326-330.
- Campbell, N. R. *An account of the principles of measurement and calculation*. London: Longmans Green, 1928.
- Centra, J. Effectiveness of student feedback in modifying college instruction. *Journal of Educational Psychology*, 1973, 65, 395-401.
- Cody, M. An application of the multiple attribute measurement model: Measurement and manipulation of source creditability. Unpublished M.A. Thesis. Michigan State University, East Lansing, Michigan. 1976.
- Cohen, A. M., Brawer, F. B. *Measuring faculty performance*. Washington, D.C.: ERIC Clearinghouse for Junior College Information/American Association of Junior Colleges. 1969.
- Costin, F., Greenough, W.T., and Menges, R. J. Student ratings of college teaching: Reliability, validity, and usefulness. *Review of Educational Research*, 1972, 41, 511-535.
- Cushman, D. P. *The message-attitude-behavior-relationship*. Bureau of Business Research, Michigan State University Press, 1977, forthcoming.
- Deshpande, A. S., Webb, S. C., Marks, E. Student perceptions of engineering instructor behaviors and their relationships to the evaluation of instructors and courses. *American Educational Research Journal*, 1970, 7, 289-305.
- Doyle, K. O. Jr. *Student evaluation of instruction*. Lexington: D.C. Heath, 1975.
- Doyle, K. O. Jr., Whitely, S. E. Student ratings as criteria for effective teaching. *American Educational Research Journal*, 1974, 11(3), 259-274.
- Eble, K. E. *The recognition and evaluation of teaching*. Salt Lake City, Utah. Project to improve college teaching. 1970.

- Einstein, A. Relativity: The special and general theory. New York: Crown. 1961.
- Fenker, R. M. Teacher evaluation at TCU: An analysis of the perceived roles of faculty by students, administrators, and faculty. In the Evaluation of College Teaching. R. I. Miller, New York: Jossey-Bass. 1974.
- Frey, P. W. Student ratings of teaching: Validity of several rating factors. Science. 1973, 182, 83-85.
- Genova, W. J., Madoff, M. K., Chin, R., and Thomas, G. B. Mutual benefit evaluation of faculty and administrators in higher education. Cambridge Mass., Ballinger Publishing Co. 1976.
- Gessner, P. Evaluation of instruction. Science, 1973, 180, 566-570.
- Gillham, J., and Woelfel, J. P. The galileo system: Preliminary evidence for precision, stability, and equivalence to traditional measures. Human Communication Research. Fall, 1976.
- Gray, C. F. The teaching model and evaluation of teaching effectiveness. Improving College and University Teaching. 1970, 18, 275-278.
- Grush, J. E., Costin, F. The student as consumer of the teaching process. American Educational Research Journal. 1975, 12, 55-66.
- Gustad, J. W. Politics and practices in faculty evaluation. Educational Record. 1961, 42, 194-211.
- Hammond, P. E., Meyer, J., Miller, D. Teaching versus research; sources of misperceptions. Journal of Higher Education. 1969, 40, 682-690.
- Hebb, D. O. A textbook of psychology. (1st ed.) Philadelphia: W. B. Saunders, 1958.
- Hebb, D. O. The American Revolution. American Psychologist. 15, 735-745, 1960.
- Helm, C. E., Messick, S., Tucker, L. Psychophysical law and scaling methods. Princeton: Educational Testing Service. 1959.
- Hildebrand, M. How to recommend promotion for a mediocre teacher without actually lying. Journal of Education. 1972. 43, 44-62.
- Hilgard, E. R., Bower, G. H. Theories of learning. New Jersey: Prentice-Hall, 1975.
- Howe, H. Less Teaching, more conversation. In C.B.T. Lee (Ed.), Improving College Teaching. Washington, D.C.: American Council on Education. 1967.

- Jacobi, C. G. J. Uber ein leichtes Verfahren die in der Theorie der Sacularstorvngen vorkommenden Gleichuagen numerisch aufzulosen. Journal fur die Reine und Angewandyte Mathematik 30:51-93.
- Kauffman, J., Hallahan, D., Payne, J., Ball, D. Teaching/Learning: Quantitative and functional analysis of education performance. Journal of Special Education. 1973, 7, 261-268.
- Keast, W. R., Macy, J. W. Jr. Faculty tenure. San Francisco: Jossey-Bass. 1973.
- Kent, L. Student evaluation of teaching. Educational Record. 1966, 47, 376-390.
- Koffka, K. Principles of gestalt psychology. New York: Harcourt-Brace. 1935.
- Kohler, W. Gestalt Psychology. New York: Liverwright. 1947.
- Lewis, L. S. Scaling the ivory tower. Baltimore Maryland: The John Hopkins University Press. 1975.
- Liska, A. E. The consistency controversy. New York: John Wiley. 1975.
- Luthans, F. The faculty promotion process and analysis of the management of large state universities. Iowa City, Iowa: Bureau of Business and Economic Research, College of Business Administration, the University of Iowa. 1967.
- McKeachie, W. J., Lin, Y., and Mann, W. Student ratings of teacher effectiveness: Validity studies. American Educational Research Journal. 1971, 8, 435-445.
- Miller, R. I. Evaluating faculty performance. San Francisco: Jossey-Bass. 1972.
- Nobel, C. An analysis of meaning. Psychological Review. 1952, 59, 421-430.
- Osgood, C. E. A behavioristic analysis of perception and language as cognitive phenomena. In Contemporary Approaches to Cognition, Cambridge: Harvard University Press, 1957.
- Osgood, C. E. An understanding and creating sentences. American Psychologist. 1963, 18, 735-751.
- Pittenger, O. E. , Gooding, C. T. Learning theories in educational practives. New York: John Wiley and Sons. 1971.
- Rippey, R. Student evaluations of professors: Are they of value? Journal of Medical Education. 1975, 50, 951-954.

- Serota, K. Metric multidimensional scaling and communication: Theory and implementation. Unpublished M.A. Thesis. East Lansing, Michigan. Michigan State University Press. 1974.
- Sharon, A. T. Eliminating bias from student ratings of college instructors. *Journal of Applied Psychology*. 1970, 54, 278-281.
- Snelbecker, B. E. Learning theory, instructional theory, and psychoeducational design. New York: McGraw-Hill. 1974.
- Stevens, S. Mathematics, measurement, and psychophysics. In S. Stevens (Ed.) *Handbook of Experimental Psychology*. New York: Wiley, 1951, 1.
- Suppes, P., Zinnes, J. L. Basic measurement theory. In R. D. Luce, R. R. Bush, and E. Galanter (eds) *Handbook of Mathematical Psychology*. New York: John Wiley and Sons. 1963, 1-76.
- Tolman, E. C. Purposive behavior in animals and men. New York: Appleton-Century-Crofts. 1932.
- Torgerson, W. Theory and methods of scaling. New York: Wiley. 1958.
- Tyler, R. W. The evaluation of teaching. In R. M. Cooper (Ed.) *The Two Sides of the Log*. Minneapolis: University of Minneapolis Press. 1958.
- Wakshlag, J. Edison, N. A model for the validation of the source creditability construct. Paper presented at the Speech Communication Association Annual Convention, December 1975.
- Wertheimer, M. Productive thinking. New York: Harper. 1959.
- Woelfel, J. Sociology and Science. Unpublished manuscript. Michigan State University. East Lansing, Michigan. 1972.
- Woelfel, J. Metric measurements of cultural processes. Paper presented to the Speech Communication Association, Chicago, 1974.
- Woelfel, J. Foundations of cognitive theory. In D. P. Cushman (Ed.), *The Message-Attitudes-Behavior Relationship*, Bureau of Business Research, Michigan State University Press, 1977, forthcoming.
- Woelfel, J., Danes, J. New techniques for the multidimensional analysis of communication and conception. In P. Monge and J. Cappella (Ed.), *Multivariate Analysis in Communication Research*, 1977, forthcoming.
- Woelfel, J., Cody, M., Gillham, J., Holmes, R. Basic premises of multidimensional attitude change theory - an experimental analysis. Unpublished manuscript, Michigan State University, East Lansing, Michigan. 1974.

Woelfel, J., Saltiel, J. Cognitive processes as motions in a multi-dimensional space: A general linear model. Unpublished manuscript. Department of Communication, Michigan State University, East Lansing, Michigan 1974.

Wotruba, T. R., Wright, P. How to develop a teacher-rating instrument. Journal of Higher Education. 1975, 46, 653-663.

Young, G., Householder, A. Discussion of a set of points in terms of their mutual distances. Psychometrika. 1938, 3, 19-22.